

Can Artificial Intelligence Itself Be Copyrighted?

By David Rabinowitz | May 12, 2025

Public discussion of artificial intelligence and copyright has focused on two main issues: training data and output.

Lawsuits have raised, and courts have begun to answer the questions (1) whether using copyrighted works as training data constitutes fair use and (2) under what circumstances artificial intelligence infringes copyright in training data.

In this article, the rules of copyright law are applied to try to answer the question whether an artificial intelligence itself might be copyrightable subject matter.

The copyrightability of computer software is not addressed. Computer software is copyrightable. *See* 17 U.S.C. §117. Artificial intelligence is the virtual machine that is generated when an artificial intelligence computer program is trained on data, resulting in a machine that applies what it has learned in response to questions or requests. More detailed definitions may be found at 15 U.S.C. §9401(3) and in the European Union’s Artificial Intelligence Act, Article 3(1).

The question that is being addressed is whether artificial intelligence can be a copyrightable work separate from the computer program that mechanically generates or operates it. That is, can artificial intelligence be copyrightable subject matter under the Copyright Act, not just the computer program that generates or operates the artificial intelligence or the computer’s memory units that store the learning from the training data, but something recognizably distinct.

The distinction for copyright purposes between a computer program and what a computer program generates has plenty of precedent. Video games and websites are good examples.

Video games are operated by computer programs, but aspects of the games – their video and audio components and some aspects of play action - are separately copyrightable. *See Upper Deck Company v. Miller*, 2024 WL 3845560 (W.D. Wash. Aug. 16, 2024) at *8 (“[s]ubstantially similar or near identical Creatures/Champions” and “unique and expressive game effects.”); *Atari Games Corp. v. Oman*, 979 F.2d

242, 245 (D.C. Cir. 1992) (R. Ginsburg, J.) (“The hallmark of a video game is the expression found in “the entire effect of the game as it appears and sounds,” its “sequence of images.” *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852, 857 (2d Cir.1982); *see also Nintendo of America, Inc. v. Elcon Indus.*, 564 F.Supp. 937, 943 (E.D.Mich.1982) (protectable “expression includes the characters, obstacles and background as well as the sequence of play of the game”); *Midway Mfg. Co. v. Bandai–America, Inc.*, 546 F.Supp. 125, 147 (D.N.J.1982) (protecting a video game’s “play and sequence of images”), *aff’d sub nom. Bandai America, Inc. v. Bally Midway Mfg. Co.*, 775 F.2d 70 (3d Cir.1985), *cert. denied*, 475 U.S. 1047, 106 S.Ct. 1265, 89 L.Ed.2d 574 (1986).”); *Spry Fox LLC v. LOLApps Inc.*, No. 2:12-CV-00147-RAJ, 2012 WL 5290158, at *5 (W.D. Wash. Sept. 18, 2012) (“A video game, much like a screenplay expressed in a film, also has elements of plot, theme, dialogue, mood, setting, pace, and character.”); although *see DaVinci Editrice S.R.L. v. ZiKo Games, LLC*, 183 F. Supp. 3d 820, 832 (S.D. Tex. 2016) (“the events in a Bang! game are not predetermined because the interactions between the roles have no underlying script or detail and are not fixed. Making certain roles aligned and others opposed is part of the game’s winning conditions, but these determine little about how players will progress through the game. *See* Boyden, 18 GEO. MASON L. REV. at 466 (copyright does not protect systems that set the stage for expression to occur).”); *see also*, Nimmer on Copyright, §2A.14[C][2] at nn. 55-60.

Artificial intelligence – the virtual machine that responds to questions and prompts – is as different from the computer that runs it as the human mind is from the neurons of the brain.

Artificial Intelligence as a Work of Authorship

The Copyright Act protects “original works of authorship fixed in any tangible medium of expression.” 17 U.S.C. §102(a). While the Copyright Act does not define “work,” it does list eight “categories” of “works of authorship.” 17 U.S.C. §102(a).

No category on the list consists of or includes artificial intelligence. But the list is not exhaustive. The legislative history of §102 says in part, “The bill does not intend either to freeze the scope of copyrightable

subject matter at the present stage of communications technology or to allow unlimited expansion into areas completely outside the present congressional intent. Section 102 implies neither that that subject matter is unlimited nor that new forms of expression within that general area of subject matter would necessarily be unprotected.” H.R. Rep. No. 94-1476, p. 51.

Is artificial intelligence a “work of authorship”? Definitions of works of authorship have been appropriately vague and expansive. The constitutional authorization to protect “writings” by copyright has been construed broadly:

“[A]lthough the word ‘writings’ might be limited to script or printed material, it may be interpreted to include *any physical rendering of the fruits of creative intellectual or aesthetic labor*.” (emphasis added) *Goldstein v. California*, 412 U.S. 546 (1973) (sound recordings).

“[A]ll forms of writing, printing, engravings, etchings, etc., *by which the ideas in the mind of the author are given visible expression*.” (emphasis added) *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884) (photographs).

A work of authorship must contain some intellectual creativity, but the quantity required is very low, far below the vast depth of an artificial intelligence: “some modest amount of intellectual labor.” *Atari Games Corp. v. Oman*, 888 F. 2d 878 (D.C. Cir. 1989) (R. Ginsburg, J.) (Pong-like game “Breakout”).

The Requirement of Human Authorship

However, a work of authorship requires human authorship.

The Copyright Office thus far has taken the position that the output of artificial intelligence is not copyrightable, on the theory that it is not the product of human creativity but rather the action of a machine. See Copyright and Artificial Intelligence, Part 2: Copyrightability, U.S. Copyright Office (Jan. 2025).

Rejecting the argument that prompts to AI can be sufficient human authorship, the Copyright Office said:

“As one commenter explained, “some element of randomness does not eliminate authorship,” but “the putative author must be able to constrain or channel the program’s processing of the source material.” The

issue is the degree of human control, rather than the predictability of the outcome.”

However, the Copyright Office has not closed the door on the possibility that human interaction with an AI might be sufficient to pass the test of human authorship:

“There may come a time when prompts can sufficiently control expressive elements in AI-generated outputs to reflect human authorship. If further advances in technology provide users with increased control over those expressive elements, a different conclusion may be called for.”

This article is concerned not with whether the Copyright Office will continue to consider AI-produced output uncopyrightable or what degree of human interaction with AI is sufficient to produce human authorship in the output, but rather whether AI itself can be a copyrightable work of authorship.

There does not appear to be any pronouncement from the Copyright Office about the copyrightability of the artificial intelligence itself, nor of the state of the computer that drives the artificial intelligence.

How would the requirement of human authorship apply to the making of an artificial intelligence?

The computer program that operates artificial intelligence is copyrightable, like any other computer program – at least the initial program before it is trained.

After the artificial intelligence program is written, the artificial intelligence is trained. Training can take different forms. Superhuman chess-playing or Go-playing computers have been trained by inputting the rules of the game, including identifying what a win consists of, and letting the computer play games against itself. Alphazero is said to have played 44 million games of chess against itself in nine hours and become a very strong player. Large language models are trained by being fed voluminous writings.

Assuming that it is human beings who select the training data, their selection of data could be copyrightable. Are the choices made in training artificial intelligences the kind of creativity that qualifies as copyrightable authorship?

Not every selection of training data would be

copyrightable. If what is selected is, for example, everything that can be collected using the Internet, there would likely be no copyrightable creativity in that selection. “Every” is not a creative or original selection. *Warren Pub. v. Microdos Data Corp.*, 115 F.3d 1509, 1518 (11th Cir. 1997) (*en banc*), *cert. den.*, 522 U.S. 963 (1997). But selections that are creative are copyrightable, even down to a low level of creative authorship. *See, e.g. Kregos v. Associated Press*, 937 F.2d 700, 704 (2d Cir. 1991) (“nine items of information concerning a pitcher’s performance” might be sufficient).

Thus, two components of artificial intelligence can each be copyrightable authorship: The computer program that operates the artificial intelligence and the selection of data that trains it. This leads to an argument that the condition of the computer that results from this combination of human authorship is itself creative authorship and therefore copyrightable.

The Copyright Office would likely reject this argument based on its position regarding artificial intelligence output. The Copyright Office requires the *expressive* elements in the output to be contributed or controlled by a human. By analogy, artificial intelligence is generated by autonomous operation of a computer program on a mass of training data and its specific elements are not dictated by humans.

Thus, it is possible that even a creative selection of the training data would not make the resulting artificial intelligence a work of human authorship. A counterargument to that is that the incorporation of the training data by the artificial intelligence happens not only according to how the data is selected, but also how the software is programmed, and the programming is a human act that incorporates human creativity.

The baking of a cake involves human intellectual activity, sometimes amounting to artistry. Every baker knows that what happens in the oven is not entirely predictable, but the baker selects the ingredients and their combination, the temperature and the time of baking. Recipes generally are not considered copyrightable, being functional and merely a description of a process or a listing of ingredients. *See* 1 Nimmer on Copyright, §2A.13. However, query whether the resulting combination of tastes, aromas and textures is any less a creative aesthetic product than a photograph.

Precedent

How have courts treated the products of human/machine interactions?

In *Burrows-Giles, supra*, the Supreme Court held that a photograph, produced by a machine, was copyrightable because of the human creativity in setting up the scene that was photographed:

“his own original mental conception, to which he gave visible form by posing the said Oscar Wilde in front of the camera, selecting and arranging the costume, draperies, and other various accessories in said photograph, arranging the subject so as to present graceful outlines, arranging and disposing the light and shade, suggesting and evoking the desired expression, and from such disposition, arrangement, or representation, made entirely by plaintiff, he produced the picture in suit.”

111 U.S. at 60.

Granted, in today’s artificial intelligence technologies, a great deal is done by the computer program in the course of training artificial intelligence that is not controlled step-by-step by the programmer in the way that a 19th Century photographer set up a picture.

Today, however, cameras contain a host of automatic features designed to improve the quality of the output and make the taking of the photograph easier and still the law continues to accept the output of modern cameras as copyrightable work of the photographer.

Similarly, in the case of an artificial intelligence, the program does much of, if not nearly all of, the work in generating the artificial intelligence, but the program is written by a programmer, and the programmer or another person chooses how to train the artificial intelligence and on which data.

How direct or specific must the connection be between human creativity and the resulting artificial intelligence, treating the artificial intelligence as the “work”?

Thomson Reuters (Westlaw) v. Ross – AI as Copyrightable Subject Matter

The idea for this article came from the case of *Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc.*, 694 F. Supp. 3d 467, 483–84 (D. Del. 2023) and a second decision in that case: 2025 WL 458520 (D. Del. 2025).

Thomson Reuters owns Westlaw, the computer-era descendant of West Publishing Company, which has been publishing judicial decisions and much other legal matter since 1872. Westlaw writes headnotes for each published decision. Westlaw's headnotes are organized by its key number system into categories and subcategories of legal issues, enabling systematic case law research.

Ross Intelligence is a new legal research company seeking to create a natural language search engine, using machine learning and artificial intelligence to avoid human intermediated materials.

If Ross is successful, users will be able to enter questions into Ross's AI search engine in ordinary English and receive back relevant quotations from judicial opinions. Ross's system would thereby find case authority directly, possibly eliminating the need for Westlaw's entire headnote and key number system.

To train its AI, Ross, through a subcontractor, created and used about 25,000 legal questions and answers. The questions were meant to be those that a lawyer would ask, and the answers were direct quotations from legal opinions.

Westlaw, however, alleges that most or all of the "created" legal questions were actually nothing more than Westlaw headnotes with a question mark appended. If true, Ross's training data could include or constitute infringing copies of Westlaw's headnotes.

One of Ross's responses was fair use. In his first decision on Westlaw's motion for summary judgment of infringement of its headnotes and Key Number system, Judge Biswas, a Third Circuit judge sitting by designation, left the fair use question for the jury:

"So whether the intermediate copying caselaw tells us that Ross's use was transformative depends on the precise nature of Ross's actions. It was transformative intermediate copying if Ross's AI only studied the language patterns in the headnotes to learn how to produce judicial opinion quotes.

But if Thomson Reuters is right that Ross used the untransformed text of headnotes **to get its AI to replicate and reproduce the creative drafting done by Westlaw's attorney-editors**, then Ross's comparisons to cases like *Sega* and *Sony* are not apt. Again, this is a material question of fact that

the jury needs to decide. (emphasis added)"

Thomson Reuters Enter. Ctr. GmbH v. Ross Intel. Inc., 694 F. Supp. 3d 467, 483–84 (D. Del. 2023).

The language quoted above is the key language in the case because of how the Ross AI operates. It accepts natural language commands or requests, but it outputs only excerpts of judicial decisions, which are public domain.

The Ross AI does not output Westlaw's headnotes. However, the court refused to dismiss Westlaw's copyright infringement claim on the first dispositive motion on the theory that "Ross used the untransformed text of headnotes to replicate and reproduce the creative drafting done by Westlaw's attorney-editors."

The "creative drafting" was the drafting that went into the array of Westlaw headnotes that Ross's AI was allegedly trained on. That is, Ross's AI was a possibly infringing copy of a body of headnotes from Westlaw's Key Number system.

If Ross's artificial intelligence can amount to an infringing copy of the "creative drafting done by Westlaw's attorney-editors" which, as the court concluded, might constitute copyright infringement, then that "creative drafting" must be copyrightable.

Presumably, that copyrightable "creative drafting" consists of the combination of the Westlaw Key Number system and the thousands of headnotes that populate that system, since this first Westlaw/Ross decision does not rest solely on the copying of the individual headnotes as training data.

If the combination of the Key Number system and its thousands of headnotes is a copyrightable work, why would the same work or a substantially similar work, if generated by and residing in an artificial intelligence, not be copyrightable subject matter?

In substance, the court posited copyright infringement by the making of an artificial intelligence, not by output from an artificial intelligence. One might say that Ross's artificial intelligence embodying Westlaw's learning was the infringing copy.

And if that learning constituted copyrightable subject matter, it would seem to follow that it would still be

copyrightable subject matter when embodied in an artificial intelligence: It is the work that is copyrightable, not the form in which it is fixed or embodied. See 1 Nimmer on Copyright, §2.03(C).

The latest decision in *Thomson Reuters v. Ross* (2025 WL 458520 (D. Del. 2025)) makes these ideas more specific. In its latest decision, the court changed its mind on Thomson Reuters' motion for summary judgment on Ross's alleged copying of 2,243 Westlaw headnotes as training data, holding that Ross had indeed infringed those headnotes. 2025 WL 458520 at *5.

More importantly for the question at hand, the court declared that the key number system itself was copyrightable:

"There is no genuine issue of material fact about the Key Number System's originality. Recall that Westlaw uses this taxonomy to organize its materials. Even if 'most of the organization decisions are made by a rote computer program and the high-level topics largely track common doctrinal topics taught as law school courses,' it still has the minimum 'spark' of originality."

The court did not go on to find the Key Number system infringed, but the key point is its copyrightability. The Key Number system, in the court's words, consists of a copyrightable "taxonomy" – a method used to "organize its materials."

Westlaw argued that "Westlaw has a registered copyright on its 'original and revised text and compilation of legal material,' which includes its headnotes and Key Number System's 'original and revised text and compilation of legal material.'" D.I. 255-7, at 8." 694 F. Supp 3d at 475. That is, Westlaw claims a copyright in its organized body of knowledge.

The court found original authorship in Westlaw's "taxonomy": "To be original, a compilation need not be 'novel,' just 'independently created by' Thomson Reuters [cit. om.]. There are many possible, logical

ways to organize legal topics by level of granularity. It is enough that Thomson Reuters chose a particular one...." 2025 WL 458520 at *3.

Summary

Westlaw's Key Number system, now recognized as copyrightable, is a body of learning that, among other things, connects lawyers with judicial decisions. What is Ross's AI but a similar body of learning that has the added feature that the user need not look for relevant headnotes, but need only put ordinary English questions to the AI?

And if Ross's AI isn't a substantial copy of the same body of learning, did Ross infringe the Key Number system by creating its AI? And whether or not Ross has infringed, such as by not copying enough of the Key Number compilation, why would Westlaw's body of learning be copyrightable in a written format but not in an AI format?

In summary, there is a reasonable case to be made that artificial intelligence can be copyrightable subject matter. Current guidance from the U.S. Copyright Office regarding artificial intelligence output seems to imply that the Copyright Office would find inadequate human authorship in the design of artificial intelligence and the selection of training materials.

It may be that the very nature of artificial intelligence, which can be said to create itself from the interaction of its computer program and training materials, will prevent the Copyright Office from registering copyright in an artificial intelligence.

However, (1) standards determining whether there is adequate human authorship of a work can evolve and (2) the courts, not the Copyright Office, have the last say in whether a work is copyrightable. It is hard to believe that the last chapter of this story has been written.

David Rabinowitz is a partner at Moses & Singer and a member of its Litigation, Intellectual Property and Law Firm Management practice groups.